

# Extraction de métriques et utilisation d'algorithmes prédictifs pour la gestion de projet

IA pour le GL - 7 Novembre 2022

---

Quentin Perez, Christelle Urtado et Sylvain Vauttier

EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Ales, France

## Génie Logiciel Empirique ? Mais encore...

Analyse des pratiques logicielles par observations et expérimentations dans le but d'en améliorer la qualité via l'adoption de nouvelles pratiques.



## Génie Logiciel Empirique ? Mais encore...

Analyse des pratiques logicielles par observations et expérimentations dans le but d'en améliorer la qualité via l'adoption de nouvelles pratiques.



- Pléthore de pratiques de développement

## Génie Logiciel Empirique ? Mais encore...

Analyse des pratiques logicielles par observations et expérimentations dans le but d'en améliorer la qualité via l'adoption de nouvelles pratiques.



- Pléthore de pratiques de développement
- Beaucoup de littérature autour du développement logiciel (parfois contradictoire)

## Génie Logiciel Empirique ? Mais encore...

Analyse des pratiques logicielles par observations et expérimentations dans le but d'en améliorer la qualité via l'adoption de nouvelles pratiques.



- Pléthore de pratiques de développement
- Beaucoup de littérature autour du développement logiciel (parfois contradictoire)
- Écart en les bonnes pratiques et la “vrai vie”
- Trouver des nouvelles méthodes de mesures de la qualité

# Utilité de l'IA dans le GL empirique

- Regrouper ou associer des données (apprentissage non-supervisé). Regrouper des sous-ensembles de projets par qualité

## Exemple

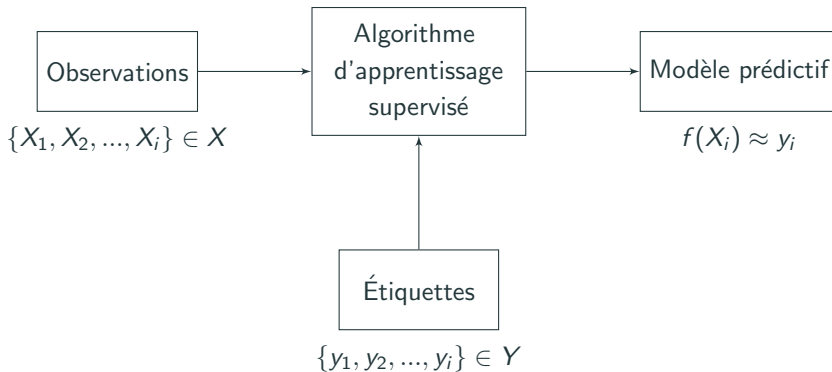
Regrouper des sous-ensembles de projets par qualité ou type de projet.

- **Classifier ou prédire des données** (apprentissage supervisé).

## Exemple

Exemple catégoriser automatiquement un ticket logiciel ou déterminer le nombre de développeurs expérimentés nécessaires en fonction de la taille du projet.

# Apprentissage supervisé



# Objectifs du TP

1. Extraire des métriques logicielles sur plusieurs versions d'un projet et en utiliser une spécifiquement (nombre de lignes de code)
2. Utiliser un classifieur (Random Forest) déjà entraîné permettant de catégoriser les développeurs avec 23 métriques sur plusieurs versions d'un projet
  - Développeurs expérimentés (*Senior Software Engineers*)
  - Développeurs non-expérimentés (*Non-Senior Software Engineers*)
3. Tracer différents graphiques
4. Créer un prédicteur du nombre de développeurs expérimentés en fonction de la taille des projets.
5. Tracer une matrice de corrélation pour étudier les différentes métriques
6. Étudier statistiquement la relation entre le fait d'être développeur expérimenté et la participation à plusieurs projets logiciels



- Langage : Python
- Extraction de données Git : GitPython
- Apprentissage machine : Scikit-Learn
- Graphiques : Matplotlib
- Extraction de métriques : CK-metrics, logiciel Java
- Environnement d'exécution : Google Colab

# Structuration du TP

TP sur Colab : <https://colab.research.google.com/drive/1dkdA1rcTDjyyMPj07ICgyVO0ND-kWjPv?usp=sharing>

Cas d'étude du TP : le projet logiciel BroadleafCommerce

## ▼ COURS-IAIL-CLASSIF-DEV

> ck\_metrics

> metrics\_by\_dev

> venv

≡ classifier\_rf.pkl

📄 dataset\_dev\_anonymized.csv

<> developers\_classification.html

📄 developers\_classification.ipynb

📄 plot\_sse\_loc\_prediction.png

📄 plot\_sse\_vs\_loc\_by\_version.png

≡ requirements.txt

- ck\_metrics : dossier contenant les métriques de BroadleafCommerce par version majeure et mineure
- metrics\_by\_dev : dossier contenant 23 métriques associées aux développeurs de BroadleafCommerce pour chaque version majeure et mineure
- classifier\_rf.pkl : classifieur Random Forest à utiliser pour catégoriser les dévs
- dataset\_dev\_anonymized.csv : fichier CSV contenant les développeurs utilisé pour entraîner le classifieur